Circumscription

Witold Łukaszewicz: Nonmonotonic reasoning: Formalization of Common-Sense Reasoning, Ellis Horwood, 1990.

Second-order logic

Alphabet:

1. A denumerable set of n-ary (for each n$\geq$0) *function variables*: $\{\phi^n_1, \phi^n_2, \ldots\}$; 0-ary function variables are called *individual variables*, and we denote them by $x_1, x_2, \ldots$.
2. A denumerable set of n-ary (for each n$\geq$0) *predicate variables*: $\{\Phi^n_1, \Phi^n_2, \ldots\}$.
3. Truth constants: $True$ and $False$.
4. Standard sentential connectives and quantifiers.
5. A countable set **P** of *predicate constants*.
6. A distinguished 2-ary predicate constant: "=".
7. A countable set **F** of *function constants*.

Terms

(1) Each individual variable is a term.

(2) Each individual constant is a term.

(3) If f is an n-ary (n>0) function constant and $\alpha_1, \ldots \alpha_n$ are terms, then $f(\alpha_1, \ldots \alpha_n)$ is a term.

(4) If ϕ is an n-ary (n>0) function variable and $\alpha_1, \ldots \alpha_n$ are terms, then $\phi(\alpha_1, \ldots \alpha_n)$ is a term.

Examples of terms:

$x,\ f(a, y),\ \phi(x, \phi(b, z))$.

Atomic formulas

(1) $True$, $False$ and all proposition constants are atomic formulas.

(2) If P is an n-ary (n>0) predicate constant and $\alpha_1, \ldots \alpha_n$ are terms, then $P(\alpha_1, \ldots \alpha_n)$ is an atomic formula.

(3) If α and β are terms, then $(\alpha = \beta)$ is an atomic formula.

(4) Each proposition variable is an atomic formula.

(5) If Φ is an n-ary (n>0) predicate variable and $\alpha_1, \ldots \alpha_n$ are terms, then $\Phi(\alpha_1, \ldots \alpha_n)$ is an atomic formula.

Examples of atomic formulas:

$P(\phi(a), a)$, $\Phi(b, x, \phi(x)), (b = \phi(y))$.

Formulas

(1) Each atomic formula is a formula.

(2) If A is a formula, then so is $(\neg A)$.

(3) If A and B are formulas, then $(A \supset B)$, $(A \wedge B)$, $(A \vee B)$ and $(A \equiv B)$ are formulas.

(4) If A is a formula and X is any variable, then $(\forall X(A))$ and $(\exists X(A))$ are formulas.

Examples of formulas:

$(\forall x \Phi(x))$, $(\forall x \exists \Psi(\neg \Psi(x) \vee \Phi(x)))$

Semantics

Definition A *frame* for a second-order language L_{AL} is a pair M=⟨D,m⟩, where D is a non-empty set, the *domain* of M, and m is a function which assigns to each n-ary (n≥0) predicate constant in AL, different from "=", an n-ary relation on D, to the constant "=", the identity relation on D, and to each n-ary (n≥0) function constant in AL, an n-ary function from D^n into D.

Given a frame M=⟨D,m⟩, we write |M| to denote the domain of M; M|K| stands for m(K), where K is any function or predicate constant in AL.

Definition An *assignment* over M is any function which assigns to every n-ary (n$\geq$0) function variable an n-ary function from $|\mathrm{M}|^n$ into |M|, and to each n-ary (n$\geq$0) predicate variable an n-ary relation on M.

The set of all assignments over M will be denoted by As(M).

If a$\in$As(M) and X is any variable, then we write $[\mathrm{a}]_X$ to denote the set of those assignments over M which differs from "a" at most on X.

Definition The *value* $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(\alpha)$ of a term α in a frame M with respect to a$\in$As(M) is an element of |M| defined as follows:

(1) $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(x) = \mathsf{a}(x)$;

(2) $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(a) = \mathsf{M}|a|$;

(3) $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(f(\alpha_1, \ldots \alpha_n)) =$

$$= \mathsf{M}|f|[\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(\alpha_1), \ldots, \mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(\alpha_n)];$$

(4) $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(\phi(\alpha_1, \ldots \alpha_n)) =$

$$= \mathsf{a}(\phi)[\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(\alpha_1), \ldots, \mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(\alpha_n)];$$

Definition The *value* $\mathsf{V}_{\mathsf{a}}^{\mathsf{M}}(A)$ of an atomic formula A in M with respect to a$\in$As(M) is an element of $\{0, 1\}$ defined as follows:

(1) $\mathsf{V}_{\mathsf{a}}^{\mathsf{M}}(True) = 1$; $\mathsf{V}_{\mathsf{a}}^{\mathsf{M}}(False) = 0$;

(2) $\mathsf{V}_{\mathsf{a}}^{\mathsf{M}}(p) = \mathsf{M}|p|$;

(3) $\mathsf{V}_{\mathsf{a}}^{\mathsf{M}}(P(\alpha_1, \ldots \alpha_n)) =$

$$= \mathsf{M}|P|[\mathsf{V}_{\mathsf{a}}^{\mathsf{M}}(\alpha_1), \ldots, \mathsf{V}_{\mathsf{a}}^{\mathsf{M}}(\alpha_n)];$$

(4) $\mathsf{V}_{\mathsf{a}}^{\mathsf{M}}(\alpha = \beta) = 1$ iff $\mathsf{V}_{\mathsf{a}}^{\mathsf{M}}(\alpha) = \mathsf{V}_{\mathsf{a}}^{\mathsf{M}}(\beta)$;

(5) $\mathsf{V}_{\mathsf{a}}^{\mathsf{M}}(\Phi^0) = \mathsf{a}(\Phi^0)$;

(6) $\mathsf{V}_{\mathsf{a}}^{\mathsf{M}}(\Phi(\alpha_1, \ldots \alpha_n)) =$

$$= \mathsf{a}(\Phi)[\mathsf{V}_{\mathsf{a}}^{\mathsf{M}}(\alpha_1), \ldots, \mathsf{V}_{\mathsf{a}}^{\mathsf{M}}(\alpha_n)].$$

Definition The *value* $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(A)$ of a non-atomic formula A in M with respect to a$\in$As(M) is an element of $\{0,1\}$ defined as follows:

(1) $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(\neg B) = 1 - \mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(B)$;

(2) $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(B \supset C) = 1$ iff $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(B) = 0$ or $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(C) = 1$;

(3) $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(B \vee C) = 1$ iff $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(B) = 1$ or $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(C) = 1$;

(4) $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(B \wedge C) = 1$ iff $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(B) = 1$ and $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(C) = 1$;

(5) $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(B \equiv C) = 1$ iff $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(A) = \mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(B)$;

(6) $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(\forall X\ B) = 1$ iff $\mathsf{V}^{\mathsf{M}}_{\mathsf{a'}}(B) = 1,$ for every a' $\in$ [a]$_X$;

(7) $\mathsf{V}^{\mathsf{M}}_{\mathsf{a}}(\exists X\ B) = 1$ iff $\mathsf{V}^{\mathsf{M}}_{\mathsf{a'}}(B) = 1,$ for some a' $\in$ [a]$_X$.

A formula A is said to be *satisfied* in M by an assignment a$\in$ As(M) iff $\mathsf{V}_{\mathsf{a}}^{\mathsf{M}}(A) = 1$.

A is *satisfiable* iff $\mathsf{V}_{\mathsf{a}}^{\mathsf{M}}(A) = 1$, for some M and some a$\in$ As(M).

A is *true* in M iff $\mathsf{V}_{\mathsf{a}}^{\mathsf{M}}(A) = 1$, for any a$\in$ As(M). In this case we say that M is a model of A.

A is valid iff A is true in every frame.

Theorem The validity problem for second-order logic is undecidable. The problem is even not semi-decidable.

A *second-order theory* is a computable set of formulas. The formulas from a theory T are called the *axioms* (*premisses*) of T.

A frame in which all the axioms of a theory T are true is called a *model* of T.

A theory T is called *satisfiable* iff it has a model.

A formula A is *entailed* by a theory T, written $T \models A$, iff A is true in every model of T.

Theories T and T' are said to be *equivalent*, written $T \Leftrightarrow T'$, iff they have the same models.

Theorem If $T = \{A_1, \ldots, A_n\}$, then

$$T \Leftrightarrow \{A_1 \wedge \cdots \wedge A_n\}.$$

THEOREM (Gödel) There is no complete deduction system for second-order logic.

Predicate and function expressions

Definition An n-ary *predicate expression* U is any expression of the form

$$\lambda x_1 \ldots x_n \; A(x_1, \ldots, x_n) \quad (n \geq 0)$$

where $x_1, \ldots, x_n$ are individual variables and $A(x_1, \ldots, x_n)$ is a formula of first- or second-order logic.

$$\boxed{U = \lambda x_1 \ldots x_n \, A(x_1, \ldots, x_n)}$$

An occurrence of X is said to be free in U iff $X \notin \{x_1, \ldots, x_n\}$ and X is free in $A(x_1, \ldots, x_n)$.

Any individual variable which is free in U is called a *parameter* in U.

If no parameters occur in U, then U is called a *predicate expression without parameters*. Otherwise U is called a *predicate expression with parameters*.

$\lambda x \; P(x); \quad \lambda xy \; [\Phi(x) \vee Q(y)]; \quad \lambda x \, Q(x, y)$

Definition Let U be a predicate expression of the form $\lambda x_1 \ldots x_n\ A(x_1, \ldots, x_n)$. The *extension* $\mathsf{E}^{\mathsf{M}}_{\mathsf{a}}(U)$ of U in a frame M with respect to a$\in$As(M) is an n-ary relation R on |M| given by

$$\langle \mathsf{d}_1, \ldots, \mathsf{d}_n \rangle \in \mathsf{R} \qquad \text{iff}$$

$$\mathsf{V}^{M}_{\mathsf{a}(x_1/\mathsf{d}_1, \ldots x_n/\mathsf{d}_n)}(A) = 1$$

where $\mathsf{a}(x_1/\mathsf{d}_1, \ldots x_n/\mathsf{d}_n)$ is the assignment on M which is like "a", except perhaps on the variables $x_1, \ldots, x_n$; here it assigns $\mathsf{d}_1, \ldots, \mathsf{d}_n$, respectively.

Note:

(1) For each n-ary predicate constant P:

$$\mathsf{E}^{\mathsf{M}}_{\mathsf{a}}[\lambda x_1 \ldots x_n \, P(x_1, \ldots, x_n)] = \mathsf{M}|P|.$$

(2) For each n-ary predicate variable Φ:

$$\mathsf{E}^{\mathsf{M}}_{\mathsf{a}}[\lambda x_1 \ldots x_n \, \Phi(x_1, \ldots, x_n)] = \mathsf{a}(\Phi).$$

$\lambda \overline{x} \; A(\overline{x})$ stands for $\lambda x_1 \ldots x_n \; A(x_1, \ldots, x_n)$.

Let $U = \lambda\overline{x}\ A(\overline{x})$, where $\overline{x} = (x_1, \ldots, x_n)$, and let $\overline{\alpha} = (\alpha_1, \ldots, \alpha_n)$ be an n-tuple of terms. The *application* of U to $\overline{\alpha}$, written $U(\overline{\alpha})$, is the formula $A(\overline{\alpha})$.

The application is *correct* iff each α_i is free for x_i in $A(\overline{x})$.

Examples

(1) If $U = \lambda xy\ P(x, y)$ and $\overline{\alpha} = (z, f(b))$, then
$U(\overline{\alpha}) = P(z, f(b))$. This is correct application.

(2) If $U = \lambda xy\ \exists z\ P(x, y, z)$ and $\overline{\alpha} = (z, f(a))$,
then $U(\overline{\alpha}) = \exists z\ P(z, f(a), z)$. This is incorect application.

A *substitution of predicate expressions for predicate variables* is a set

$$\{\Phi_1 \leftarrow U_1, \ldots, \Phi_n \leftarrow U_n\} \qquad (n \geq 0)$$

where $\Phi_1, \ldots, \Phi_n$ are distinct predicate variables, $U_1, \ldots, U_n$ are predicate expressions and the corresponding Φ_i and U_i have the same arities.

Definition The *application* of a substitution $\theta = \{\Phi_1 \leftarrow U_1, \ldots, \Phi_n \leftarrow U_n\}$ to a formula A, written $A\theta$, is the formula obtained from A by the following construction:

(1) Replace each free occurrence of Φ_i in A by $\lfloor U_i \rfloor$.

(2) In the expression resulted from (1), which in general is not a formula, successively replace each part of the form $\lfloor U_i \rfloor(\alpha)$ by the application of U_i to α.

The application of θ to A is *proper* iff

(i) All the applications (of predicate expressions to terms) performed in (2) are correct.

(ii) No free occurrence of any variable in any of $U_1, \ldots, U_n$ becomes bound in $A\theta$.

Example Let

$A = \forall x \Phi(x) \supset \forall x \Phi(b); \quad \theta = \{\Phi \leftarrow \lambda y P(x, y)\}.$

Applying θ to A, we first get

$$\forall x \lfloor \lambda y\ P(x, y) \rfloor (x) \supset \forall x \lfloor \lambda y\ P(x, y) \rfloor (b)$$

and then

$$\forall x\ P(x, x) \supset \forall x\ P(x, b).$$

Example Let

$A = \forall x \Phi(x) \supset \Phi(y); \quad \theta = \{\Phi \leftarrow \lambda x \exists y Q(x, y)\}.$

Applying θ to A, we first get

$$\forall x \, \lfloor \lambda x \, \exists y \, Q(x, y) \rfloor (x) \supset \lfloor \lambda x \, \exists y \, Q(x, y) \rfloor (y)$$

and then

$$\forall x \, \exists y \, Q(x, y) \supset \exists y \, Q(y, y).$$

This is an improper application because the application of $\lambda x \ \exists y \ Q(x, y)$ to y is not correct (y is not free for x in $\exists y \, Q(x, y)$).

Example Let

$$A = \forall xy(\Phi(x,y) \vee \Psi(x,y))$$

$$\theta = \{\Phi \leftarrow \lambda xy P(x,y), \Psi \leftarrow \lambda uv Q(u,v)\}.$$

Applying θ to A we first obtain

$$\forall xy(\lfloor \lambda xy P(x,y) \rfloor (x,y) \vee \lfloor \lambda uv Q(u,v) \rfloor (x,y))$$

and then

$$\forall xy(P(x,y) \vee Q(x,y)).$$

This is a proper application.

Theorem For any formula A and any substitution $\theta = \{\Phi_1 \leftarrow U_1, \ldots, \Phi_n \leftarrow U_n\}$, the formula

$$(\forall \Phi_1 \ldots \Phi_n A) \supset A\theta$$

is valid, provided that the application of θ to A is proper.

Corollary For any sentence of the form $\forall \Phi_1 \ldots \Phi_n A$ and any substitution $\theta = \{\Phi_1 \leftarrow U_1, \ldots, \Phi_n \leftarrow U_n\}$,

$$\forall \Phi_1 \ldots \Phi_n A \models A\theta$$

provided that the application of θ to A is proper.

Definition An n-ary *function expression* F is any expression of the form

$$\lambda x_1 \ldots x_n \, \alpha(x_1, \ldots, x_n) \quad (n \geq 0)$$

where $x_1, \ldots, x_n$ are individual variables and $\alpha(x_1, \ldots, x_n)$ is a term.

The following are function expressions:

$$\lambda x f(x); \quad \lambda x y \; f(g(x), y); \quad \lambda x \phi(x, y).$$

An n-ary function expression is meant to represent an n-ary function.

Circumscription

We distinguish between:

- a predicate, thought of as a particular relation specified over a particular domain, and
- predicate expressions, including predicate constants and predicate variables, which can be used to denote this relation.

Recall:

- An n-ary predicate constant P can be identified with the predicate expression $\lambda x_1 \ldots x_n \, P(x_1 \ldots x_n)$.
- An n-ary predicate variable Φ can be identified with the predicate expression $\lambda x_1 \ldots x_n \, \Phi(x_1 \ldots x_n)$.
- The predicate denoted by a predicate expression U is called the *extension* of U.

General idea of circumscription

The objects (tuples of objects) that can be shown to satisfy a certain predicate are the **only** objects (tuples of objects) enjoying it.

More specifically: Given a theory T and an n-ary predicate constant P, to circumscribe P in T is to assume that the extension of P is minimal in the sense that it cannot be made smaller without contradicting T.

Circumscription is thus a form of minimization: to circumscribe a predicate constant is to minimize its extension.

Let T consist of

$$Red(a) \wedge On(a, b).$$

Circumscribing Red in T, we expect to infer

$$\forall x \; Red(x) \supset x = a.$$

Circumscribing On in T, we expect to derive

$$\forall xy. \; On(x, y) \supset x = a \wedge y = b.$$

Circumscription can be also used to minimize extensions of several predicates.

All versions of circumscription share the following common characteristics:

- Unlike other non-monotonic formalisms, circumscription allows us to formalize non-monotonic reasoning directly in the language of classical first-order logic.

- It is always the task of the user to specify predicates which are to be minimized. Circumscription provides a general method which can be applied to arbitrarily chosen predicates.

- Circumscription is based on syntactic manipulations. Given a theory T and a list of predicate constants $P_1, \ldots P_n$, circumscription of $P_1, \ldots P_n$ in T amounts to implicitly adding a special second-order sentence (a set of first-order formulas, in some versions of circumscription) capturing the desired minimization.

Abnormality formalism

(1) How the common-sense knowledge, including non-monotonic rules, is to be represented in the framework of circumscription?

(2) Which predicates are to be minimized to assure intuitively acceptable conclusions?

$\boxed{Ab \ - \ \text{abnormal}}$

$$\forall x.\, Bird(x) \wedge \neg Ab(x) \supset Flies(x)$$

(1) $Bird(Tweety)$

(2) $\forall x.\, Bird(x) \wedge \neg Ab(x) \supset Flies(x).$

Conclusions:

$\forall x \neg Ab(x)$

$\forall x.\, Bird(x) \supset Flies(x)$

$Flies(Tweety).$

(3) $Bird(Clyde) \wedge \neg Flies(Clyde)$

Conclusions:

$\forall x.\, x \neq Clyde \supset \neg Ab(x)$

$\forall x.\, Bird(x) \wedge x \neq Clyde \supset Flies(x).$

(1) $Bird(Tweety)$

(2) $\forall x.\, Bird(x) \wedge \neg Ab(x) \supset Flies(x)$

(3) $Penguin(Joe)$

(4) $\forall x.\, Penguin(x) \supset Bird(x).$

Conclusion:

$Flies(Joe).$

How to block this conclusion?

$\forall x.\, Penguin(x) \supset \neg Flies(x).$

$\forall x.\, Penguin(x) \supset Ab(x).$

$$\forall x.\ Penguin(x) \wedge \neg Ab(x) \supset \neg Flies(x).$$

This is incorrect representation.

$$\forall x.\ Penguin(x) \wedge \neg Ab_1(x) \supset \neg Flies(x).$$

Ab and Ab_1 should be jointly minimized.
Unfortunately, there is a conflict.

Minimization$_1$: $\neg Ab(Joe)$ and $Flies(Joe)$

Minimization$_2$: $\neg Ab_1(Joe)$ and $\neg Flies(Joe)$.

$$\forall x.\ Penguin(x) \supset Ab(x).$$

There are non-monotonic rules requiring many-place abnormality predicates:

> "In the absence of evidence to the contrary, assume that x gives y a gift, provided that y has a birthday and x is a friend of y" .

$$\forall xy.\ Birthday(y) \wedge Friend(x,y) \wedge \neg Ab(x,y) \supset Gives\text{-}Gift(x,y).$$

Here $Ab(x,y)$ should be read: "x behaves abnormally with respect to y in the situation where y has a birthday and x is a friend of y" .

Predicate circumscription (McCarthy 1980)

L - a fixed first-order language with equality.

The objects under consideration are **finite** sets of sentences from L, referred to as *theories*.

Recall that every such a theory can be identified with a single sentence, namely with the conjunction of all its members.

We write $T(P_1, \ldots P_n)$ to indicate that some (but not necessarily all) of predicate constants occurring in T are among $P_1, \ldots P_n$.

L_2 - a second-order language whose alphabet is that of L, together with an infinite set of n-place (n$\geq$0) predicate variables.

Let U and V be n-ary predicate expressions. We write

$$U \leq V \quad \text{for} \quad \forall \overline{x}.\, U(\overline{x}) \supset V(\overline{x})$$

where $\overline{x} = (x_1, \ldots x_n)$ is a new n-tuple of individual variables and $\forall \overline{x}$ stands for $\forall x_1 \cdots \forall x_n$.

Note: $U \leq V$ states that the extension of U is a subset (not necessarily proper) of the extension of V.

Let $\overline{U}$ and $\overline{V}$, where $\overline{U} = (U_1, \ldots, U_n)$ and $\overline{V} = (V_1, \ldots, V_n)$, be similar tuples of predicate expressions, i.e. U_i and V_i have the same arities, for each $1 \leq i \leq n$. We write $\overline{U} \leq \overline{V}$ to denote the sentence

$$\bigwedge_{i=1}^{n} [U_i \leq V_i] \quad \text{i.e.} \quad \bigwedge_{i=1}^{n} [\forall \overline{x}.\, U_i(\overline{x}) \supset V_i(\overline{x})].$$

$\overline{U} \leq \overline{V}$ states that the extension of each member of $\overline{U}$ is a subset (not necessarily proper) of the extension of the corresponding member of $\overline{V}$.

Let $T(P_1, \ldots, P_n)$ be a theory. How to express the fact that the extensions of $P_1, \ldots, P_n$ are minimal with respect to T, i.e. none of them can be made smaller without contradicting T?

$$\forall \Phi_1 \ldots \Phi_n . \left[T(\Phi_1, \ldots, \Phi_n) \wedge \bigwedge_{i=1}^{n} (\Phi_i \leq P_i) \right] \supset \bigwedge_{i=1}^{n} (P_i \leq \Phi_i).$$

Here $\Phi_1, \ldots, \Phi_n$ are predicate variables with the same arities as $P_1, \ldots, P_n$, respectively. $T(\Phi_1, \ldots, \Phi_n)$ is the sentence obtained from $T(P_1, \ldots, P_n)$ by replacing all occurrences of $P_1, \ldots, P_n$ by $\Phi_1, \ldots, \Phi_n$, respectively.

Notice: The presence of $T(\Phi_1, \ldots, \Phi_n)$ in the above formula assures that the extensions of $P_1, \ldots, P_n$ are minimized relative to T.

Definition The *circumscription schema* of $P_1, \ldots, P_n$ in T, denoted by CS(T; $P_1, \ldots, P_n$), is the second-order formula

$$\left[T(\Phi_1, \ldots, \Phi_n) \wedge \bigwedge_{i=1}^{n} [\Phi_i \leq P_i]\right] \supset \bigwedge_{i=1}^{n} [P_i \leq \Phi_i]$$

or, equivalently

$$\left[T(\Phi_1, \ldots, \Phi_n) \wedge \bigwedge_{i=1}^{n} [\forall \overline{x}.\ \Phi_i(\overline{x}) \supset P_i(\overline{x})]\right] \supset$$

$$\supset \bigwedge_{i=1}^{n} [\forall \overline{x}.\ P_i(\overline{x}) \supset \Phi_i(\overline{x})]$$

where $\Phi_1, \ldots, \Phi_n$ and $T(\Phi_1, \ldots, \Phi_n)$ are defined as before.

Example Let $T = \{Red(a) \wedge On(a,b)\}$.

$\mathsf{CS}(T; Red) =$

$$[\Phi(a) \wedge On(a,b) \wedge [\forall x.\ \Phi(x) \supset Red(x)]] \supset$$

$$\supset [\forall x.\ Red(x) \supset \Phi(x)].$$

Definition Let $T(P_1, \ldots, P_n)$ be a theory and let $\overline{\mathsf{CS}}(T; P_1, \ldots, P_n)$ be the set of all formulas obtainable from $\mathsf{CS}(T; P_1, \ldots, P_n)$ by proper substitutions of first-order predicate expressions for $\Phi_1, \ldots, \Phi_n$. The *predicate circumscription* of $\overline{P} = (P_1, \ldots, P_n)$ in T, written $\mathsf{CIRC}_{\mathsf{PR}}(T; \overline{P})$, is $T \cup \overline{\mathsf{CS}}(T; P_1, \ldots, P_n)$.

The set of theorems derivable by circumscribing $\overline{P} = (P_1, \ldots, P_n)$ in T is identified with the set of theorems classically derivable from $\mathsf{CIRC}_{\mathsf{PR}}(T; \overline{P})$.

Example Let $T = \{Red(a) \wedge On(a,b)\}$.

$\mathsf{CS}(T; Red) =$

$$[\Phi(a) \wedge On(a,b) \wedge [\forall x.\ \Phi(x) \supset Red(x)]] \supset [\forall x.\ Red(x) \supset \Phi(x)].$$

We show that

$$\mathsf{CIRC}_{\mathsf{PR}}(T; Red) \models \forall x.\ Red(x) \supset x = a \quad (1)$$

Substituting $\lambda x\ x = a$ for Φ, we get

$$[a = a \wedge On(a,b) \wedge [\forall x.\ x = a \supset Red(x)]] \supset$$
$$\supset [\forall x.\ Red(x) \supset x = a].$$

From this, we immediately conclude (1).

Example We circumscribe "=" in $T = \{P(a)\}$.

$\mathsf{CS}(T; =) =$

$$[P(a) \wedge [\forall xy.\, \Phi(x, y) \supset x = y]] \supset [\forall xy.\, x = y \supset \Phi(x, y)].$$

Putting $\{\Phi \leftarrow \lambda xy\, False\}$, we get

$$[P(a) \wedge [\forall xy.\, False \supset x = y]] \supset [\forall xy.\, x = y \supset False]$$

which is equivalent to $\qquad P(a) \supset \forall xy.\, x \neq y.$

Since $T \vdash P(a)$, we finally conclude that $\mathsf{CIRC}_{\mathsf{PR}}(T; =)$ is inconsistent. But if we add equality axioms the above derivation is blocked.

Semantics for predicate circumscription

Definition Let $T(P_1, \ldots, P_n)$ be a theory, some of whose predicate constants are among $\overline{P} = (P_1, \ldots, P_n)$. A model M of T is a $\overline{P}$-*submodel* of a model N of T, written M$\leq^{\overline{P}}$N, iff

(1) |M| = |N|;

(2) M$|P| \subseteq$ N$|P|$, for every predicate constant P in $\overline{P}$;

(3) M$|K|$ = N$|K|$, where K is any function constant or any predicate constant not in $\overline{P}$.

M is a $\overline{P}$-*minimal* model of T iff every model of T which is a $\overline{P}$-submodel of M is identical with M.

Theorem If $\mathrm{CIRC}_{\mathsf{PR}}(T; \overline{P}) \vdash A$, then A is true in every $\overline{P}$-minimal model of T.

Example Let T consist of:

$Bird(Tweety)$

$\forall x.\, Bird(x) \wedge \neg Ab(x) \supset Flies(x)$.

Consider the frame M, where:

$|\mathsf{M}|=\{\textsc{tweety}\}$; $\mathsf{M}|Flies|=\{\}$;

$\mathsf{M}|Bird|=\mathsf{M}|Ab|=\{\textsc{tweety}\}$.

M is an Ab-minimal model of T (we cannot make $\mathsf{M}|Ab|$ smaller without contradicting T). Thus, since $Flies(Tweety)$ is false in M, this formula cannot be derived by circumscribing Ab in T.

This example illustrates a more general phenomenon:

THEOREM: Neither positive nor negative new ground instances of uncircumscribed predicates can be derived by predicate circumscription.

In view of the above theorem, **predicate circumscription is generally incapable to deal with non-monotonic rules!**

Reason: While minimizing the extensions of circumscribed predicates, predicate circumscription keeps the extensions of all other predicates fixed.

Second-order circumscription (McCarthy, Lifschitz)

Second-order circumscription extends predicate circumscription in three ways:

1. The process of minimization is expressed by a special *circumscription axiom*, being a single second-order sentence rather than a set of first-order formulas.

2. The extensions of arbitrarily chosen predicate constants, also of those not minimized, may be varied during the minimization.

3. The extensions of arbitrarily chosen function constants may be also allowed to vary in the process of minimization.

Recall:

(1) If U and V are predicate expressions of the same arity, then $U \leq V$ stands for $\forall \overline{x}.\, U(\overline{x}) \supset V(\overline{x})$.

(2) If $\overline{U} = (U_1, \ldots, U_n)$ and $\overline{V} = (V_1, \ldots, V_n)$ are similar tuples of predicate expressions, then $\overline{U} \leq \overline{V}$ stands for

$$\bigwedge_{i=1}^{n} [U_i \leq V_i] \quad \text{i.e.} \quad \bigwedge_{i=1}^{n} [\forall \overline{x}.\, U_i(\overline{x}) \supset V_i(\overline{x})].$$

(3) We write $\overline{U} = \overline{V}$ for $(\overline{U} \leq \overline{V}) \wedge (\overline{V} \leq \overline{U})$, and $\overline{U} < \overline{V}$ for $(\overline{U} \leq \overline{V}) \wedge \neg(\overline{V} \leq \overline{U})$.

Definition Let $\overline{P}$ be a tuple of distinct predicate constants, let $\overline{S}$ be a tuple of distinct function and/or predicate constants disjoint with $\overline{P}$, and suppose that $T(\overline{P}, \overline{S})$ is a first- or second-order theory. The *second-order circumscription of* $\overline{P}$ in T *with variable* $\overline{S}$, written $\mathrm{CIRC}_{\mathsf{SO}}(T; \overline{P}; \overline{S})$, is the sentence

$$T(\overline{P}, \overline{S}) \wedge \forall \overline{\Phi\Psi}\, \neg[T(\overline{\Phi}, \overline{\Psi}) \wedge \overline{\Phi} < \overline{P}]$$

where $\overline{\Phi}$ and $\overline{\Psi}$ are tuples of variables similar to $\overline{P}$ and $\overline{S}$, respectively.

A formula A is said to be a *consequence* of second-order circumscription of $\overline{P}$ in T with *variable* $\overline{S}$ iff

$$\mathrm{CIRC}_{\mathsf{SO}}(T; \overline{P}; \overline{S}) \models A.$$

$\mathsf{CIRC}_{\mathsf{SO}}(T; \overline{P}; \overline{S})$ can be rewritten as

$$T(\overline{P}, \overline{S}) \wedge \forall \overline{\Phi\Psi} \left(T(\overline{\Phi}, \overline{\Psi}) \wedge (\overline{\Phi} \leq \overline{P}) \supset (\overline{P} \leq \overline{\Phi}) \right)$$

which, in turn, is an abbreviation for

$$T(\overline{P}, \overline{S}) \wedge \forall \overline{\Phi\Psi} \left[\left(T(\overline{\Phi}, \overline{\Psi}) \wedge \bigwedge_{i=1}^{n} [\forall \overline{x}.\ \Phi_i(\overline{x}) \supset P_i(\overline{x})] \right) \supset \right.$$

$$\left. \supset \bigwedge_{i=1}^{n} [\forall \overline{x}.\ P_i(\overline{x}) \supset \Phi_i(\overline{x})] \right].$$

Example Let $T = \{Bird(Tweety) \wedge$

$$\wedge\, [\forall x.\, Bird(x) \wedge \neg Ab(x) \supset Flies(x)]\}.$$

We take $\overline{P} = (Ab)$ and $\overline{S} = (Flies)$.

$$\mathsf{CIRC}_{\mathsf{SO}}(T; \overline{P}; \overline{S}) = T(\overline{P}, \overline{S}) \wedge \forall \Phi \Psi$$

$$\Big(Bird(Tweety) \wedge [\forall x.\, Bird(x) \wedge \neg \Phi(x) \supset \Psi(x)] \wedge$$

$$\wedge\, [\forall x.\, \Phi(x) \supset Ab(x)]\Big) \supset \forall x.\, Ab(x) \supset \Phi(x).$$

Taking $\{\Phi \leftarrow \lambda x False; \Psi \leftarrow \lambda x Bird(x)\}$, we get:

$$\mathsf{CIRC}_{\mathsf{SO}}(T; \overline{P}; \overline{S}) \models \forall x.\, Ab(x) \supset False$$

$$\mathsf{CIRC}_{\mathsf{SO}}(T; \overline{P}; \overline{S}) \models \forall x.\, \neg Ab(x)$$

$$\mathsf{CIRC}_{\mathsf{SO}}(T; \overline{P}; \overline{S}) \models \forall x.\, Bird(x) \supset Flies(x)$$

$$\mathsf{CIRC}_{\mathsf{SO}}(T; \overline{P}; \overline{S}) \models\ \ Flies(Tweety).$$

Semantics for second-order circumscription

Theorem Let $\overline{P}$, $\overline{S}$ and $T(\overline{P}, \overline{S})$ be as before. Let M and N be models of T. M is a $(\overline{P}; \overline{S})$-*submodel* of N, written $\mathsf{M} \leq^{(\overline{P};\overline{S})} \mathsf{N}$, iff

(1) $|\mathsf{M}| = |\mathsf{N}|$;

(2) $\mathsf{M}|P| \subseteq \mathsf{N}|P|$, for every predicate constant P in $\overline{P}$;

(3) $\mathsf{M}|K| = \mathsf{N}|K|$, for any function or predicate constant K not in $\overline{P}, \overline{S}$.

We write $\mathsf{M} <^{(\overline{P};\overline{S})} \mathsf{N}$ iff $\mathsf{M} \leq^{(\overline{P};\overline{S})} \mathsf{N}$ but not $\mathsf{N} \leq^{(\overline{P};\overline{S})} \mathsf{M}$. A model M of T is $(\overline{P}; \overline{S})$-*minimal* iff T has no model N such that $\mathsf{N} <^{(\overline{P};\overline{S})} \mathsf{M}$.

MOD$^{(\overline{P};\overline{S})}(T)$ - the class of all $(\overline{P};\overline{S})$- minimal models of T.

MOD(T) - the class of all models of T.

Theorem For any T, $\overline{P}$ and $\overline{S}$,

MOD(CIRC$_{\mathsf{SO}}(T;\overline{P};\overline{S})$)=MOD$^{(\overline{P};\overline{S})}(T)$.

Satisfiability of second-order circumscription

Theorem Second-order circumscription need not preserve satisfiability.

Definition A theory T is *well-founded with respect to* $(\overline{P};\overline{S})$ iff, for every M$\in$MOD(T), there is N$\in$ MOD$^{(\overline{P};\overline{S})}(T)$ such that N$\leq^{(\overline{P};\overline{S})}$ M.

Theorem If T is satisfiable and well-founded with respect to $(\overline{P};\overline{S})$, then CIRC$_{\mathsf{SO}}(T;\overline{P};\overline{S})$ is satisfiable.

Well-foundedness is not necessary for satisfiability of second-order circumscription.

A sentence is said to be *universal* iff its prenex normal form contains no existential quantifiers. A theory is said to be *universal* iff all of its axioms are universal.

Theorem If T is universal, then T is well-founded with respect to any $(\overline{P}; ())$.

Corollary For any tuple of predicate constants $\overline{P}$ and any satisfiable universal theory T, $\mathrm{CIRC}_{\mathrm{SO}}(T; \overline{P}; ())$ is satisfiable.

Theorem Second-order circumscription need not preserve satisfiability for universal theories if function constants are allowed to vary.

The situation is different if varying constants are restricted to predicate constants:

Theorem For any disjoint tuples $\overline{P}$ and $\overline{S}$ of predicate constants, every universal theory is well-founded with respect to $(\overline{P}; \overline{S})$.

Corollary For any disjoint tuples $\overline{P}$ and $\overline{S}$ of predicate constants and any satisfiable universal theory T, $\mathrm{CIRC}_{\mathsf{SO}}(T; \overline{P}; \overline{S})$ is satisfiable.

Expressive power of second-order circumscription

Theorem If $T(\overline{P}, \overline{S})$ is well-founded with respect to $(\overline{P}; \overline{S})$, $P \in \overline{P}$ is an n-ary predicate constant, $\overline{S}$ is a tuple of predicate constants disjoint with $\overline{P}$, and $\overline{\alpha}$ is an n-tuple of ground terms, then

$$\mathrm{CIRC}_{\mathsf{SO}}(T; \overline{P}; \overline{S}) \models P(\overline{\alpha}) \text{ iff } T \models P(\overline{\alpha}).$$

Theorem If $T(\overline{P}, \overline{S})$ is well-founded with respect to $(\overline{P}; \overline{S})$, $\overline{P}$ and $\overline{S}$ are disjoint tuples of predicate constants, R is an n-ary predicate constant not in $\overline{P}, \overline{S}$, and $\overline{\alpha}$ is an n-tuple of ground terms, then

(i) $\mathrm{CIRC}_{\mathsf{SO}}(T; \overline{P}; \overline{S}) \models R(\overline{\alpha})$ iff $T \models R(\overline{\alpha})$
(ii) $\mathrm{CIRC}_{\mathsf{SO}}(T; \overline{P}; \overline{S}) \models \neg R(\overline{\alpha})$ iff $T \models \neg R(\overline{\alpha})$.

Note: The only ground literals that may be obtained from well-founded theories by using second-order circumscription (without varying function constants) are negative instances of circumscribed predicates and instances, both positive and negative, of variable predicates.

This is not true if function constants are allowed to vary. In such a case, both **positive literals of circumscribed predicates** and **new instances of non-variable predicates** may be obtainable.

No new formulas can be obtained by circumscribing "=":

Theorem For any T, $\overline{S}$ and A:

$$\mathrm{CIRC}_{\mathrm{SO}}(T; =; \overline{S}) \models A \quad \text{iff} \quad T \models A.$$

The *domain closure assumption*, with respect to a theory T, is the assumption that the only individuals to be considered are those forced by the Herbrand universe of T. In the case where T gives rise to a finite Herbrand universe $\{c_1, \ldots, c_n\}$, the domain closure assumption can be formalized by employing the *domain closure axiom* (Reiter):

$$\forall x.\, x = c_1 \vee \cdots \vee x = c_n.$$

Second-order circumscription cannot be used, at least for well-founded theories and without varying functions, to obtain the domain closure axioms:

Theorem If $T(\overline{P}, \overline{S})$ is well-founded with respect to $(\overline{P}; \overline{S})$, $\overline{S}$ contains no function constants, and $\alpha_1, \ldots, \alpha_n$ are ground terms, then

$$\mathrm{CIRC}_{\mathsf{SO}}(T; \overline{P}; \overline{S}) \models \forall x.\, x = \alpha_1 \vee \cdots \vee x = \alpha_n$$

$$\text{iff} \quad T \models \forall x.\, x = \alpha_1 \vee \cdots \vee x = \alpha_n.$$

Frequently, second-order circumscription axiom can be replaced by an equivalent first-order sentence. This allows to avoid second-order logic and makes the formalism very attractive for the purpose of practical applications (see Łukaszewicz: "Non-Monotonic Reasoning" pp. 249-252).